

## Durham Research Online

---

### Deposited in DRO:

24 November 2016

### Version of attached file:

Accepted Version

### Peer-review status of attached file:

Peer-reviewed

### Citation for published item:

See, B.H. and Gorard, S. and Siddiqui, N. (2017) 'Can explicit teaching of knowledge improve reading attainment? An evaluation of the core knowledge curriculum.', *British educational research journal*, 43 (2). pp. 372-393.

### Further information on publisher's website:

<https://doi.org/10.1002/berj.3278>

### Publisher's copyright statement:

This is the accepted version of the following article: See, B.H., Gorard, S. Siddiqui, N. (2017). Can explicit teaching of knowledge improve reading attainment? An evaluation of the core knowledge curriculum. *British Educational Research Journal* 43(2): 372-393, which has been published in final form at <https://doi.org/10.1002/berj.3278>. This article may be used for non-commercial purposes in accordance With Wiley Terms and Conditions for self-archiving.

## Use policy

---

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

# Can explicit teaching of knowledge improve reading attainment?: An evaluation of the core knowledge curriculum

Beng Huat See, Stephen Gorard and Nadia Siddiqui  
Durham University  
b.h.see@durham.ac.uk

## Abstract

In England, as elsewhere, there is a tension in primary schools between imparting knowledge and teaching basic skills like literacy and numeracy. State-mandated programmes are generally concerned with structure and skills. However, a number of ministers and advisers across administrations have sought to expand the explicit teaching of world knowledge (Culture, Geography, and Science) as advocated by E.D. Hirsch in the Core Knowledge curriculum. This paper describes an independent evaluation of an adaptation of that approach, called the 'Word and World Reading programme' as used with children aged 7 to 9 in England, to assess its impact on wider literacy. Nine primary schools were randomised to receive the intervention from the start and another eight a year later. The outcomes, assessed by the Progress in English test in literacy after one year, showed no discernible effect overall ('effect' size -0.03), and a small improvement for those eligible for free school meals (+0.06). There was no school dropout, but the missing data for around 18% of 1,628 pupils means the results must be treated with some caution. Observations suggest that the lack of clear benefits could be due to the poor quality of implementation in some schools or classes. Perhaps teachers as professionals do not respond well to prescriptive curricula. It is also possible that factual knowledge does not translate directly to improved literacy skills, at least not within one year. Teaching children facts alone in this way cannot be justified solely in terms of improved literacy. Even the scheme on which this intervention was based stressed the need for pupils to learn how to handle facts as well as to learn the facts themselves.

## Introduction

There are attempts in the US and now the UK to teach wider and greater content knowledge to primary age children. A theory behind this movement is that children, especially from disadvantaged backgrounds, do not read well partly because they do not possess the necessary background knowledge to make sense of what they read. One programme addressing this is the Core Knowledge Language Arts curriculum (CKLA) whose aim is to expose children to new words and concepts so that the new words stay in their long-term memory and thus facilitate future learning. And if children understand the words they read they can understand the text. In cases where children have less exposure to a wide range of vocabulary, they do not have the background knowledge to build on or a context in which to place what they are trying to read. Their learning is therefore hindered, so that the gap between good and poor readers widens.

Core Knowledge Language Arts (CKLA) is a programme developed by E.D. Hirsch. It first gained popularity in the US after the publication of the book *Cultural Literacy* (Hirsch 1987). According to Hirsch, literacy depends to a great extent on understanding of context and textual references and so on the possession of relevant facts. Hirsch developed his idea when he observed that some groups of students were able to understand passages of texts more easily than others, and that this systematic difference was due to lack of familiarity with the context (Hirsch, 1987). He believed that this was because the curriculum in the early years did not allow children to build on their basic background knowledge. To help develop children's comprehension skills Hirsch and his foundation developed the Core Knowledge Language Arts curriculum. The idea was to expose children to informational texts and literary non-fiction to build their vocabulary and generic background knowledge.

The programme was first piloted in 1990 in a Florida elementary school. Since then it has gone through several revisions based on feedback from schools, and has been implemented in thousands of schools all over the US.

In recent years the core knowledge curriculum has received more attention in England. Several influential commentators in England including a minister for schools (Nick Gibb) and an Education Secretary (Michael Gove) have spoken openly of their admiration of Hirsch's philosophy (The Guardian 2012). In 2013, Michael Gove set out an agenda to reform the curriculum to put emphasis on the teaching of core knowledge (Coughlan 2013). Two new primary schools were set up in London by the think tank Civitas specifically using a curriculum that is built on the philosophy of CKLA. The journalist Toby Young also opened a secondary free school in West London basing its curriculum on the Core Knowledge Sequence. Early indications are that these schools are popular with parents and places are oversubscribed. However, the impact of CKLA on academic attainment has not been evaluated. As the programme gained popularity in England and elsewhere it is therefore timely to evaluate its impact, especially given that policy-makers and stakeholders are increasingly calling for policies and practice to be informed by research evidence.

To test the effectiveness of the Core Knowledge programme in UK, The Curriculum Centre (TCC) applied for funding from the Education Endowment Foundation (EEF) to trial a pilot of their version of the programme, known as Word and World Reading (WWR). This is the programme described in this paper. TCC is part of Future Academies, whose Director (Lord Nash) was Parliamentary Under-Secretary for Schools. We were appointed by the EEF as independent evaluators to assess the impact of WWR. We had no input into the programme, and were concerned only with whether it held promise to improve literacy in UK schools.

WWR takes its inspiration from the Core Knowledge Language Arts curriculum but also draws on the work of other academics such as Walter Kintsch. Its main developer, Daisy Christodoulou, is the author of the much talked about book 'Seven myths about education' (Christodoulou 2014). In the book she puts forth the principles and theory behind the core knowledge curriculum drawing on evidence from research in cognitive science.

To investigate these claims further, this paper looks at the existing evidence, and then describes the methods and findings of a mixed-method evaluation of a specific

version of CKLA. The paper ends by considering the implications for policies, schools and future studies.

## **Prior evidence**

The key feature of the Core Knowledge curriculum is an emphasis on content and information rather than skills and methods. This has sparked a debate over the need to teach with more emphasis on content as opposed to skills. Many commentators have disagreed with Hirsch's views that learning content knowledge is the way towards cultural literacy. Tchudi (1988), for example, argued that cultural literacy cannot be achieved by piling facts on children; it is acquired through participation and living in society. Booth (1988) disagreed with Hirsch's idea that children acquire knowledge by being taught knowledge. Educators need to understand the processes by which children assimilate such knowledge. Schwarz (1988) is highly critical of what she sees as Hirsch's dismissive attitude towards critical thinking. She criticised his vision as being narrow, elitist and content-specific, and suggested that Hirsch's idea to revamp a whole curriculum without taking into account how the knowledge should be taught and the nature of the knowledge itself was a sure way to bring about a decline in cultural literacy. Proponents of the thinking skills philosophy believed that the foundation to literacy is the ability to question, reason and argue. Critiques of Hirsch's theory also argued that a content-focussed curriculum is problematic because the content is prescriptive and is therefore very reliant on the views of those who dictated it (Squire 1988).

So, what is the existing evidence of the impact of the core knowledge curriculum on literacy? And is there any evidence that teaching critical thinking or reasoning skills is any better for this purpose? Throughout readers must recall that both content and reasoning can be useful contributors to literacy and both are likely to be valuable for their own sake (Gorard et al. 2016).

Prior evidence for the impact of the CK curriculum on literacy has so far been weak. A longitudinal study following 301 seventh and eight grade children over three years using a matched comparison design reported mixed effects (Whitehurst-Hall 1999). This study measured impact using the Iowa Test of Basic Skills (ITBS) subtests on reading, language and maths. Positive impacts were noted for some measures but not others. The authors reported no 'statistically significant' differences between CK and control pupils in terms of grade retention and grade failure.

The biggest study conducted across four states in the US also showed mixed outcomes. Norm-referenced tests were used to compare four matched pairs of CK and comparator schools over three years (Stringfield et al. 2000, and see follow up by Dartnow et al. 2000). The overall impact of CK on reading was negative both for children who started at aged six (effect size of -0.06) and those who received the programme at aged 8 (effect size of -0.08). The programme appeared to be more beneficial for low achieving younger pupils (effect size of +0.25), but not for older pupils (effect size -0.53). Schools in all of the states (Washington, Maryland and Texas) apart from Florida registered a negative impact on reading. The authors explained that the exceptionally poor performance of CK pupils in the low implementing school in Maryland had skewed the overall results.

The Core Knowledge website (Core Knowledge Foundation 2000) reported an independent evaluation across schools in Oklahoma city. It suggested that on average CK pupils performed better than comparison pupils in reading comprehension, vocabulary and social studies on the ITBS Norm-Referenced test (ITBS) and the Oklahoma Criterion-Referenced Tests. However, the full report was not available and the summary extract highlighted only the positives, and did not give any indication of attrition. One study suggested that the Core Knowledge curriculum might be more effective for very young children. This was a pilot study of the CK programme in early years setting (The NYC Core Knowledge Early Literacy Pilot 2012). It reported promising gains in reading tests (Woodcock-Johnson III) especially in kindergarten, although the differences decreased by the third year. Using the standardised TerraNova test, however, no differences were detected in oral reading comprehension and vocabulary. Although the report suggested that children who were on the programme for the longest had the highest post-test scores compared to those with only one and two years of exposure, these children had higher pre-test scores too. In other words, they started from a higher base score.

There are some concerns that the prescriptive nature of the Core Knowledge sequence might stifle creativity. To address such concerns Baer (2011) compared the creative writing (poems and short stories) of seventh and eighth Grade pupils (n=540) in Core Knowledge schools with those in non-Core Knowledge schools. The writing exercises were graded independently by experienced writers and teachers using the Amabile (1982) Consensual Assessment Technique. The results showed that 7<sup>th</sup> Grade CK pupils outperformed non-CK pupils while 8<sup>th</sup> Grade CK pupils did worse.

So far therefore, the evidence has been inconclusive - positive effects were reported for only some year groups, in some states and for only some sub-tests. When taken together the overall results appeared to be negative in almost all but one study conducted in one city in the US. All the large-scale evaluations on the academic impact of the CKLA programme tended to use matched comparison designs, which is not ideal. Pupils were not randomly assigned to the intervention. This means that differences in outcomes may be due to some unobserved differences between pupils. And even if pupils were precisely matched, once there is attrition the groups become unequal. Most studies did not report the levels of attrition.

Critiques of Hirsch's philosophy (as mentioned above) suggest that teaching content knowledge itself is not enough. Children should also be equipped with the skills to extract, evaluate and retain information. Thinking skills, for example, are just as relevant where children learn how to use information and ideas effectively, to critique, reason and argue. One intervention attempting to do this is Philosophy for Children (P4C). The evidence for this approach appears to be more positive, in terms of literacy outcomes, than the core knowledge curriculum. A systematic review of evidence for P4C conducted by Trickey and Topping (2004) showed consistent moderate effects on a range of outcome measures. The mean 'effect' size for the studies included was 0.43. P4C was also shown to have suggested positive long term effects on literacy in a experimental study by Topping and Trickey (2007). A large randomised controlled trial of P4C (Gorard et al. 2017) conducted in the UK with primary school children reported promising results for reading (ES = 0.12) and maths (0.10), but negligible differences for writing (ES = 0.03). These were measured as

progress from KS1 to KS2. The programme was found to be more effective for children eligible for free school meals (ES = + 0.29 for reading; ES = +0.17 for writing and ES = +0.20 for maths). Several high quality studies have also shown the positive effects of thinking skills on a range of subjects (e.g. Worth et al. 2015; Hanley et al. 2015; Reznitskaya 2012).

Others have also argued that it is not what is taught but how it is taught that mattered. There is some observational evidence that the quality of teaching practice may be associated with student achievement (Cantrell et al. 2008). It is generally agreed that to be effective teachers, be it in maths, science or literacy, a good grounding in the subject knowledge is necessary. And equally important is the skill to convey that knowledge to learners in a way that can be easily assimilated (Gorard et al. 2016).

### **Aims of the new study**

The main aim of our new pilot trial was to assess the promise of the Word and World Reading programme, based on a curriculum focused on the teaching of factual knowledge, for improving reading comprehension and wider literacy skills of primary school pupils. The programme emphasized the teaching of content knowledge. The role of the teacher was simply to teach that content as the protocol dictates. The quality of teaching and the effectiveness of the teachers are all meant to be controlled through a highly prescriptive text. Teachers are given the teaching materials and instructed how to read the text to children, what questions to ask, how to ask them and at which juncture to pose the questions. There is very limited leeway for teachers to exercise their individuality. In this way it would help to test the evidence base of a programme that is solely reliant on the transmission of knowledge. As this was the first ever RCT of the Core Knowledge curriculum, a secondary aim was to assess the intervention itself, the resources and teaching materials and teachers' skills in delivering such a knowledge-intensive curriculum. This paper is *not* about curriculum development nor is it about development of knowledge or curricula in general. It is specifically to test Hirsch's idea that simply transmitting knowledge to children using a prescriptive curriculum to teach facts can improve literacy skills of young children. The focus is therefore on the impact of the intervention although we also examined the process of implementing the programme as it provides information on how the intervention might be replicated in their specific context, or why the programme did not achieve the impact intended if found to be ineffective, and if successful, what are the mechanisms that lead to change. Although it does not contribute to new knowledge as such, this paper seeks to provide conclusive answer as to whether teaching facts alone is enough to develop literacy skills in children as proponents of the programme suggested. Is it as simple as that or is there more to it? If found to be successful, it could transform the way children learn and develop literacy skills.

### **The intervention**

The programme being evaluated is known as Word and World Reading (WWR), developed by The Curriculum Centre (TCC). The WWR programme takes its inspiration from the US Core Knowledge Language Arts (CKLA) programme using the same Core Knowledge Sequence. The core curriculum being taught in this study involved knowledge of world History and Geography. The aim of the teaching was to

expose children to new words and new concepts repeatedly over the course of a year so that the new words would stay in their long-term memory and thus facilitate future learning. The theory is that if children understand the words they read they can understand the text. In cases where children have less exposure to a wide range of vocabulary, they do not have the background knowledge to build on. The WWR programme is to provide this opportunity to children to help build their vocabulary to assist them with their reading. The fundamental theory behind this approach to learning is reportedly based on cognitive science about how memory works.

The Word and World Reading programme is a whole class intervention, carried out twice per week over one school year as part of the literacy class. Each lesson lasts 45 minutes, and is taught by literacy teachers. There are 34 pre-planned Geography and 35 History lessons. The lessons are very structured, following a set sequence. Every lesson begins with a reading passage, no more than a page long. The teacher reads the passage aloud and then pauses and asks questions about the text. A few new words are introduced explicitly in every session. There is a lot of repetition of keywords/concepts. To reinforce the learning of these keywords pupils answer a few short mastery questions in their workbook followed by a keyword exercise. One aspect of the intervention is instant feedback. Teachers are expected to go round to check pupils' answers and mark their workbook, giving immediate feedback or pencil in some suggestions for improvement. As part of the methodology pupils are required to use full sentences. The emphasis is on acquisition of knowledge and using the keywords correctly.

The teaching resources include textbooks for pupils, teacher handbooks, globes and atlases. The pupil textbook is organised in units and each unit consists of a sequence of passages that are linked, allowing pupils to build their conceptual understanding and vocabulary. The texts are written in a simple language appropriate for the age group. The textbook is also the bound workbook.

The teacher handbook is similar to the pupil textbook, but with guidance notes for teachers and prompts indicating where to stop and repeat the information for students or engage them in discussion. Suggested questions are included. Some suggestions are provided on the use of images. Every teacher is given a globe and every class has seven globes to be shared among the pupils. Additionally, there is also an atlas for every child. All Year 3 and Year 4 literacy teachers in the treatment schools attended a one-day training learning to use the sequence as described above, and the materials and activities that came with it.

The intervention was delivered in the experimental schools, while the control schools carried on their literacy lessons as normal during the trial.

### **Conducting the impact evaluation**

The trial was conducted using a standard two-group waitlist design. This meant that one group of schools received the intervention immediately and the other group received the intervention a year later after the trial was completed. The advantage of this is that it incentivised control schools to remain in the trial, reducing

demoralisation and thus dropout. It is also ethical, meaning that no school was deprived of the resources available.

### *Sample*

The schools involved in this trial were 17 primary schools (1,628 pupils) recruited from a range of geographical areas across England by The Curriculum Centre (the programme developer) through their contacts. These included schools of different sizes (five-form entry to one-form entry) in areas of high social deprivation, inner city schools with ‘high challenge’ pupils, and schools serving predominantly white working class communities. To minimise the possibility of cross-contamination the schools recruited were those not already involved in other similar programmes. Pupils eligible for the trial were all those in Years 3 and 4 initially (age 7 to 9). The intention was to recruit a wide range of schools with different intakes, mixed ethnicity and different levels of disadvantage across England. The schools included seven academy converters, nine academy sponsored schools and one community school.

### *Randomisation*

Nine schools were randomised to the treatment group and the other eight to the waitlist control. Randomisation was carried out in front of colleagues to maintain objectivity and transparency. School level randomisation was used because the programme was delivered as a whole-class intervention across two year groups. Although class and year group randomisations were considered, school level randomisation was deemed to carry the least risk of diffusion.

Once recruitment was confirmed the pre-test was administered to all the Year 3 and Year 4 pupils in the schools. Randomisation was undertaken after the pre-test to ensure blinding, because knowledge of group allocation could affect pupils’ performance in the test and teachers’ attitude towards the test. For this reason also, the administration of the post-test was monitored by the evaluators since schools would now know which group they are in. Evaluators sent invigilators to participating schools to ensure that the test was taken under exam conditions, that schools took the test seriously and that it was carried out consistently and fairly between groups.

The groups were found to be slightly unbalanced at the outset (Table 1) in terms of background characteristics. There were substantially more ethnic ‘minority’ pupils and those for whom English is not their first language in the treatment group. Presumably this imbalance was because of the small number of cases randomised (17 schools).

Table 1 – Pupil characteristics in treatment and control groups at the beginning of the trial

| Pupil characteristic | Treatment | Control   |
|----------------------|-----------|-----------|
| Male                 | 322 (49%) | 350 (51%) |
| FSM-eligible         | 237 (36%) | 183 (27%) |
| SEN                  | 112 (17%) | 171 (25%) |
| EAL                  | 298 (45%) | 164 (24%) |
| Non-White British    | 505 (77%) | 308 (45%) |



### *Attrition*

Over the course of the trial, nine pupils were withdrawn across all schools. These were pupils identified with special educational needs and were given another intervention deemed more appropriate by their schools, and in contradiction to the agreement for the trial. In addition, a large number of pupils from one treatment school did not take the post-test. This accounted for 43% of the attrition. Some teachers mistakenly thought that since they did not intend to continue with the programme the following year, they did not have to do the post-test. Schools also sometimes excluded pupils with learning difficulties and those with severe behavioural and emotional difficulties from tests. Teachers explained that they thought these children would not be able to access the test and this would put undue stress on them. This explained why proportionately more boys and SEN pupils did not take the post-test. A total of 287 post-test results were missing (206 from the treatment schools, and 81 from control schools), representing an overall attrition rate of 18%. This is most likely to bias the findings in favour of the intervention group, because it tended to be the weaker pupils who dropped out and more dropped out from the treatment group.

### *Outcome measures and analyses*

The impact of the intervention was measured by comparing the gain scores between the pre and post-test of both groups using the long version of the Progress in English (PiE) test. PiE is a standardised test of reading and comprehension.

The difference in gains between groups was converted to Hedge's  $g$  effect size, calculated as the difference between the gains scores of the treatment group and the control group divided by the pooled standard deviation. Sub-group analyses were also conducted to see if the programme was particularly beneficial for pupils eligible for free school meals (FSM).

To help illustrate how much of the difference in scores can be attributed to the WWR programme, we also conducted a 2-step multivariate regression analysis using the gain scores as the dependent variable, and the available context variables as predictors in the first step, and whether each pupil was in the treatment group or not in the second step.

We have intentionally not used significance tests and confidence intervals in the analyses for a number of reasons. One is that they are not relevant because significance tests do not tell us what we really want to know, that is, whether the data shows that there is a 'real' difference between groups (Carver 1978; Watts 1991) and it can lead to serious mistakes (Falk and Greenbaum 1995, Gorard 2016). Significance tests are also predicated on complete random samples with no attrition (Lipsey et al. 2012; Berk and Freedman 2001). As noted above, attrition was 18% and the missing data/cases were clearly not random in occurrence. Pupils with missing post-test data were selected out, incorrectly by schools, because they were deemed unable to take the test. Missing cases were also likely to include long term sick, permanently excluded, transient population from migrant community and "school refusers" (pupils who refused to attend schools).

To assess the security of the finding, that is whether it could occur by chance or due to bias as a result of attrition, we calculated the ‘number needed to disturb the finding’ of the effect size (NNTD). This is the number of counterfactual cases that would be needed to alter the substantive findings. In other words it would take that number of cases to change the results. The bigger the number the more stable the finding is. The counterfactual score is derived in the following steps:

- Calculate the mean of the gain scores or post-test scores (if comparison is with the post-test scores) of the two groups.
- If the mean of the larger group is greater than that of the smaller group, then we add its standard deviation to its mean. If it is smaller we subtract the standard deviation from its mean. In this study, the control group is larger than the treatment group and its mean is greater (see Table 2), so we add the standard deviation of the control group to its mean.
- Then we add the counterfactual scores so created as to the smaller group (in this case it is the treatment group), and continue adding until the difference between the group means disappears or reverses.
- This is the number of cases needed to erase any difference between groups. The number of such scores that can be added to the treatment group before the effect size disappears is the ‘number needed to disturb the finding’ (Gorard and Gorard 2016). The control counterfactual is used here because the treatment group is slightly smaller (and so is easier to ‘disturb’).

This evaluation used intention-to-treat analyses. This means that all eligible pupils who were randomised (including those who had left school) were included in the analyses as far as possible. Therefore, all efforts were made to track these pupils. Despite this, many leavers could not be traced or their destination schools were unable to cooperate.

## **Conducting the process evaluation**

The purpose of the process evaluation was to assess the fidelity (that is if the programme was delivered as intended) and quality of implementation of the WWR intervention (Carroll et al. 2007; Montgomery et al. 2013). This also helped provide an explanation for the mechanism (or theory of change) of how the intervention works, if it is successful. And if the programme was found to be ineffective it could explain whether the programme is intrinsically ineffective, or the teachers were not implementing it in the way they should. It also helps to identify potential limitations (Steckler and Linnan 2002) and capture unintended consequences (side effects) if any.

The process evaluation was designed as an adjunct to the impact evaluation. To avoid the results of the impact evaluation influencing what we reported in our process evaluation, the process evaluation report was written independently of the impact evaluation.

The process evaluation largely involved classroom observations of the delivery of lessons and interviews with pupils and staff. Classroom observation visits to schools were arranged with the lead programme trainer. These were carried out once at the beginning of the intervention to observe the delivery of the programme, noting

inconsistencies or any departures from the programme protocol and also to note pupils' reaction to the programme and teachers' ability to use the resources. Another round of visits was carried out towards the end of the intervention, this time looking for changes in teachers' behaviour and any differences in children's learning. There was no formal observation schedule as such, but all researchers were given a standard briefing with an outline of what to look for, and to make notes of all processes in the lessons. The broad briefing ensures that we capture all aspects of the intervention in practice and are not limited or constrained by any preconceived ideas. We have intentionally kept it broad so that we are not influenced by what we think is important, but what the participants think is important, and to allow the evidence to speak for itself without prejudice. So, all information is potentially relevant here.

During the school visits researchers would sit in the lesson as participant observers. They would also walk around the class from table to table to look at pupils' work and listen to their comments. There were often opportunities to talk to pupils while they were doing the writing activities. All observations, pupils' feedback and remarks were noted, but not digitally recorded.

In total 12 visits were made to 8 treatment schools to observe the process of implementation in 56 lessons. Group interviews with teachers and pupils were carried out during these visits.

Assessments of the quality, relevance and utilisation of teaching resources (teachers' handbooks and pupil's workbook, the pictorial images, globes and atlases), and staff and pupils' perceptions of these materials also formed part of the process evaluation.

The interviews were informal and focused on assessing teachers' and pupils' perceptions of the programme regarding what they thought had contributed or would contribute to the success of the programme, and the barriers to effective delivery of the intervention.

Group interviews were held with teachers, averaging four to five teachers each. We also interviewed three heads and one deputy head. Interviews were held with research lead teachers in all the schools we visited, and we asked teachers:

- What they liked about the programme?
- Whether they have observed any impact on children's learning?
- What challenges they faced?
- How can the programme be improved?

Two group interviews were conducted in each of the schools, one with Y3 and one with Y4 pupils. In some schools pupils from both year groups were interviewed together. On average the groups consisted of six children, and we asked:

- What they liked about the programme?
- What they did not like about the programme?
- How can the programme be improved?
- Whether there was any impact on their learning?

In addition, we also made sample visits to eight schools (3 control and 5 intervention schools) to observe the administration of the post-test to see if the testing was carried out fairly and consistently across schools. Some of the test visits also included lesson observations and interviews.

After each visit, observation reports were written up and interview data were transcribed and collated according to the themes of the questions above and in line with the aims of the process evaluation (see p.9 &10). All interview responses were handwritten as schools did not routinely allow interviews to be recorded. Some new themes emerged as a result of our observations, and these were also included in the analysis because we did not want to allow our questions to constrain the kind of evidence that might emerge. Essentially, our analysis focused on questions that were specific to the intervention (e.g. teaching materials) and issues related to the implementation of the intervention. We were particularly interested in whether the programme was implemented as intended and factors that facilitate or hinder the effective implementation of the programme. This is one of the aims of an efficacy trial – to identify successful features and areas where the intervention can be improved in order to make recommendations for future roll out of the programme, as well as to provide answers to what makes the intervention works (if found to be successful) and why the intervention fails (if found to be ineffective). So the process evaluation was analysed with these questions in mind and the results presented accordingly.

## Results of the impact evaluation

The results based on the gain scores showed no discernible benefit for reading comprehension measured using the PiE tests (effect size = -0.03; Table 2). The result was the same using standardised age scores instead of raw scores. The number of counterfactual cases that would be needed to disturb the finding would be 17 (Gorard and Gorard 2016), which is much less than the level of attrition, suggesting that there was no clear impact either way.

Table 2 – Analysis using raw scores, comparing gain scores for all schools

|         | N    | PiE pre-test | Standard deviation | PiE post-test | Standard deviation | Gain score | Standard deviation | ‘Effect’ size |
|---------|------|--------------|--------------------|---------------|--------------------|------------|--------------------|---------------|
| TCC     | 659  | 22.9         | 8.4                | 47.5          | 15.7               | 24.7       | 12.6               | -0.03         |
| Control | 678  | 22.2         | 8.9                | 47.8          | 16.3               | 25.1       | 12.9               | -             |
| Total   | 1337 | 22.6         | 8.6                | 47.7          | 16.0               | 24.9       | 12.7               | -             |

*Note: One control school provided only post-test scores. Therefore the gain scores have N of 565 in the control group.*

The one control school that did not take the pre-test had a slightly higher post-test score (50.6; SD14.2) than the overall average of 47.7 (Table 2). So if we include the post-test scores for this school we may artificially reduce the apparent effect size of the intervention. To check if this was the case, we excluded this school from the analysis. The overall finding remained the same (effect size = -0.03; Table 3). This means that including this school in the analysis will make no difference to the substantive outcome.

Table 3 – Analysis using raw scores for all schools, excluding the school without pre-test scores

|         | N    | PiE pre-test | Standard deviation | PiE post-test | Standard deviation | Gain score | Standard deviation | ‘Effect’ size |
|---------|------|--------------|--------------------|---------------|--------------------|------------|--------------------|---------------|
| TCC     | 659  | 22.9         | 8.4                | 47.5          | 15.7               | 24.7       | 12.6               | -0.03         |
| Control | 565  | 22.2         | 8.9                | 47.3          | 16.6               | 25.1       | 12.9               | -             |
| Total   | 1224 | 22.6         | 8.6                | 47.4          | 16.1               | 24.9       | 12.7               | -             |

To see if the programme impacted on older and younger children differently, we analysed the two year groups separately. The results were similar for both groups (Tables 4 and 5), suggesting that the programme had no particular benefit for either year group.

Table 4 – Analysis using raw scores for all schools, Year 3 only

|         | N   | PiE pre-test | Standard deviation | PiE post-test | Standard deviation | Gain score | Standard deviation | ‘Effect’ size |
|---------|-----|--------------|--------------------|---------------|--------------------|------------|--------------------|---------------|
| TCC     | 304 | 21.6         | 8.5                | 46.4          | 14.5               | 24.8       | 11.2               | -0.03         |
| Control | 307 | 21.6         | 8.7                | 46.9          | 15.7               | 25.2       | 12.5               | -             |
| Total   | 611 | 21.6         | 8.6                | 46.7          | 15.2               | 25.0       | 11.8               | -             |

Table 5 - Analysis using raw scores for all schools, Year 4 only

|         | N   | PiE pre-test | Standard deviation | PiE post-test | Standard deviation | Gain score | Standard deviation | ‘Effect’ size |
|---------|-----|--------------|--------------------|---------------|--------------------|------------|--------------------|---------------|
| TCC     | 355 | 24.0         | 8.1                | 48.5          | 16.5               | 24.5       | 13.7               | -0.04         |
| Control | 258 | 22.9         | 9.0                | 48.9          | 16.9               | 25.0       | 13.4               | -             |
| Total   | 613 | 23.5         | 8.5                | 48.7          | 16.7               | 24.7       | 13.6               | -             |

We also wanted to see if the programme benefited poorer children (pupils eligible for FSM) in particular. Our analysis suggested that the programme may have had a slight positive effect for these children (effect size of +0.06, Table 6). However, because these children were not randomly allocated and because of the relatively small number of FSM pupils, we have to be careful about reading too much into this result. The number of counterfactual cases needed to disturb this result is only 9, suggesting that there is no clear impact.

Table 6 - Raw score results, all schools, FSM-eligible pupils only

|         | N   | PiE pre-test | Standard deviation | PiE post-test | Standard deviation | Gain score | Standard deviation | ‘Effect’ size |
|---------|-----|--------------|--------------------|---------------|--------------------|------------|--------------------|---------------|
| TCC     | 237 | 22.9         | 8.6                | 46.2          | 15.9               | 23.3       | 12.5               | +0.06         |
| Control | 157 | 20.4         | 9.0                | 42.8          | 16.3               | 22.5       | 14.2               | -             |
| Total   | 394 | 21.9         | 8.9                | 44.7          | 16.1               | 23.0       | 13.2               | -             |

To help contextualise how much of any difference in outcomes might be explained by participation in the WWR programme, we conducted multivariate analyses using two outcomes (gain scores and post-test scores). Pupil background variables and their pre-test scores were entered in Step 1, and whether pupils were in the treatment group or not in Step 2. The results of this analysis displayed in Table 7 show that 42% of the difference in post-test scores can be ‘explained’ by factors prior to the intervention ( $R=0.65$ ,  $R^2=0.42$ ). This model seems to be better at explaining differences in the post-test scores than the gain scores ( $R=0.27$ ,  $R^2=0.08$ ). Knowing whether a pupil was in the treatment group or not added little to the difference in either gain or post-test scores. Although this was not a test of causation it does confirm the finding that the programme made little difference to pupils’ outcomes, even when differences in the initial pupil characteristics are taken into account.

Table 7 - Variation explained (R) in two-stage regression model, using two possible outcomes

|  | Gain score outcome | Post-test outcome |
|--|--------------------|-------------------|
| Step 1 – background and prior attainment | 0.27               | 0.65              |
| Step 2 – intervention                    | 0.28               | 0.65              |

When we considered all of the explanatory variables, the pupil’s age at pre- and post-test and their pre-test scores (for gain score outcomes) are the best predictors of their performance in reading as measured by the PiE test (Table 8). Knowing whether the child was in the programme or not does not help to predict their performance. This suggests that the intervention had little to do with how pupils performed on the standardized test.

Table 8 - Standardised coefficients for the regression model in Table 11

|                            | Gain score outcome | Post-test outcome |
|----------------------------|--------------------|-------------------|
| FSM                        | +0.01              | -0.09             |
| Sex (female)               | +0.08              | -0.05             |
| SEN                        | +0.06              | -0.06             |
| EAL                        | +0.01              | -0.10             |
| Ethnicity (White UK)       | -0.06              | -0.15             |
| Age at pre-test            | -1.04              | -0.30             |
| Age at post-test           | +0.98              | +0.30             |
| PiE (pre-test)             | -0.56              | +0.06             |
| Step 2: Treatment (or not) | -0.03              | +0.07             |

## Results from the process evaluation

The results of the process evaluation are not definitive as it is based on observations of sample lessons and participants’ self-reports. However, it is important in that it provides clues to help explain the results of the impact evaluation. The observations and comments reported here represent a general picture. We do not generally quantify the number of occurrences nor name the schools (for reasons of anonymity) as the point is not about how many or who said what. The

examples given are merely to illustrate the points we want to make. For example, there were many examples of poor teaching across all the schools (see below), but we only give a few examples to demonstrate the difficulties in introducing such a programme effectively in UK schools.

*Was there any perceived impact on pupils' learning?*

Although the impact evaluation (above) showed no obvious advantage from the WWR programme on the reading comprehension skills of primary school pupils, teachers reported evidence of change in some aspects of pupils' learning.

Some teachers reported a perceptible impact on children's writing. Others said the practice of answering in full sentences was useful for other tests. The habit had cascaded out to other work. Teachers reported students using sentence starters, which they had applied in other subjects. Some also spoke about how students had become more confident in using technical terms. Others felt that the programme had helped hone pupils' comprehension skills. A deputy head spoke about how pupils often excitedly told her what they had been learning in the lessons. She felt that the intervention had "contributed significantly".

There was some evidence that pupils were thinking about what they had learnt during the lessons. For example, one pupil asked the evaluator whether it was proper to use "death" or "died", such as "after his death" or "after he died". However, it is difficult to say if this was the effect of the intervention. What is clear is that pupils were using the keywords learnt in their writing. For example, terms associated with the compass points (N, S, E, W), seasons, trench and cities, farming, invention and environment were evident in their writing. One headteacher told us that they could see beneficial effects on children's vocabulary.

It's extensive recap going over a number of terms helps the class with their confusions - over the difference between "civilisation" and "citizen" and between "emperor" and "empire".

Further evidence of possible learning could be seen in pupils' workbook exercises. For example, pupils demonstrated their grasp of quite abstract concepts such as 'democracy' and 'government' in the sentences they made. One child wrote about how they had difficulty deciding what to do, so the group had a vote and made the decision democratically, illustrating the pupil's understanding of the word, 'democratic'. However, these instances were rarely observed and there was no evidence that the child would not be able to use such a vocabulary anyway without the intervention.

Changes in other aspects of learning were also observed. Some teachers noted that pupils were listening better and working a lot faster. One teacher noticed an obvious difference between a newly arrived pupil and the other pupils, but said that the child soon began to catch up, having been exposed to the programme.

From the pupils' perspectives, they said that what they learnt in WWR helped in other lessons, particularly the International Primary Curriculum (IPC). For example, they also learnt about volcanoes and civilisations in IPC.

Informal chats with pupils suggested that they believed that the programme had an impact on their learning. A Year 4 pupil thought the extended questions helped her to develop ideas:

You can go back to the text and read, not just write with your head down. You can write your own things. It makes us think a lot. There is a lot of “what do you think”. It’s our opinions that we write.

Pupils also said that the programme had also helped them with “big-writing”, that it had helped them to write more. A Year 4 pupil said:

It gets us used to comprehension in tests. It’s really good practice for writing in full sentences so we don’t lose marks when we’re doing tests.

Pupils also mentioned enjoying learning about new places and one said that he wanted to travel and the lessons would help “when you travel to new places”.

*What factors could have hindered the effective implementation of the programme?*

One of the aims of the project was to pilot the curriculum. The process evaluation identified aspects of it which could potentially limit the effectiveness of the programme.

#### *Lack of differentiation in the curriculum*

In this study, the developers decided to use the same syllabus for both year groups with the same topics, texts and exercises. Teachers reported that some of the topics and concepts were beyond the grasp of the younger pupils.

There was also a lack of differentiation by ability. For example, pupils with special educational needs and those for whom English was an additional language (EAL) were reported to find the topics challenging. One teacher remarked that some of her Romanian EAL pupils could not be included in the lessons at all. All teachers we spoke to expressed concern that the lack of differentiation might disadvantage those working below the expected level. In a number of lessons we observed we noticed that the less able pupils struggled with the activities in the workbook. Many were distracted and some simply abandoned the writing task. The only differentiation was the additional extension activity in the exercise book for the more able pupils. This exercise consisted of an open-ended question, often requiring pupils to think more broadly about the topic. Feedback from the pupils was that these activities were ‘samey’ with little variation, and not challenging enough in this formulation.

#### *Inadequate supporting teaching materials*

As part of the teaching package teachers were provided with pictures to supplement the lessons. A common remark from teachers was that there were insufficient pictures they could use to support their lessons. Some of the pictures were also not of very good quality – too small and unclear.



### Unappealing texts

The children's textbook was not very inspirational. It was simply photocopied texts in black and white and put together in flimsy ring binders. These came apart after a week of use. There were no coloured images, no pictures or diagrams in the book. The children we spoke to often asked for more colourful images in their textbooks.

### Lack of opportunity to engage in in-depth discussions

Some teachers found it hard to adapt the highly prescriptive structure of the curriculum to their teaching style. As a result lessons appeared forced and contrived. There was little attempt by teachers to engage in discussions and develop any themes further. This could be due to a lack of confidence to venture beyond the set text or inadequate background knowledge, but presumably mainly because teachers were told to adhere strictly to the protocol.

Feedback from teachers suggested that they would have liked to have the opportunity to discuss some topics in greater detail. Pupils also told us that they would have liked to have less coverage of the curriculum, and more discussions about each topic. They said the lessons sometimes seemed 'rushed'.

Although there were some excellent lessons where the teachers brought in their own experience and generated interesting discussions, this was not the case in a majority of the lessons observed. And where teachers did try to expand on the content of the lessons it was not always successful. One teacher, for example, introduced one session by reminding children about a lesson on Rome that they had learnt in another class. She started off by saying that Julius Caesar was the emperor of Rome only to be contradicted when she read the text. This illustrates that for teachers to be able to engage in discussions they need to have the requisite knowledge.

In one lesson on ancient civilisations, children asked how people knew what it was like in those days since none of us alive today was around then. The teacher ignored the question and moved on. In a geography class the teacher explained currents as movement of water under the surface of the sea. In an earlier lesson they were told that waves were movement of water on the surface of the sea caused by wind. The children then asked how water under the surface of the sea moved because there was no wind under the sea. Again this was not taken up for further discussions. There were ample opportunities like these where teachers could have developed the lessons further with more interesting examples and explanations. These opportunities were missed and students' interests were not engaged. In a lesson on irrigation the teachers explained irrigation as ditches that carry water and that ditches were lines in the ground. The teacher could not respond when children asked how digging ditches helped move water.

There were ample opportunities like these where teachers could have developed the lessons further with more interesting examples and explanations. These opportunities were missed and students' interests were not engaged.

### Teachers' level of relevant knowledge

Teachers' lack of relevant knowledge in the subjects they are teaching could be a potential hindrance to the successful implementation of the programme. This was observed in lessons across all of the schools. One teacher actually said they did not

used to know anything about Geography and History. We got the sense that most of the literacy teachers were learning about the subjects as they went along. This problem was also noted in the first national study of the Core Knowledge Curriculum in the US (Stringfield et al. (2000, p. 30). For example, one Principal in the study commented that *‘some teachers initially lacked the background knowledge in specific content of the Core Knowledge Sequence.’* Another Principal said, ‘teachers look at the curriculum and say, ‘I don’t know anything about this, and I don’t have time to learn it,’ and they brush it aside.”

There were many instances where teachers made factual errors. These occurred in the majority of lessons observed. One teacher confused latitude with longitude. Another could not distinguish between a sand dune coast and a rocky coast. One teacher explained that ‘tropical’ meant hot and wet when describing the tropical rainforest, but later described the tropical savannah as hot and dry. The children were understandably confused as they had just learnt that ‘tropical ‘ meant hot and wet.

In *one* lesson, a number of factual and pronunciation errors were noted. This lesson is used as an example because it illustrates the kind of mistakes that teachers could make in teaching content subjects.

- The teacher talked about the Mariana Trench as the ‘Marina Trench’ and accepted a pupil’s answer that a trench was the deepest part of the ocean. The Mariana Trench may be the deepest part of the ocean but a trench is not.
- The teacher mistakenly thought coral reefs were living things. This is inaccurate as ‘corals’ are living things, but ‘reefs’ are not.
- The teacher told pupils whales are fish and that they have a sprout (sic) rather than a spout at the back of their neck (rather than the top of the head)
- The teacher confused whales with shark and told the children

Sometimes the questions teachers asked also did not make sense. For example, ‘What do you call the bottom of the sea floor?’

In a History lesson the teacher wanted to teach children the concept of cities, and using Great Yarmouth (GY) as an example, told children there were no horses or farms in cities. Children had seen horses and farms in GY, so were not convinced that GY was a city.

In another History lesson, when asked to name some things that Mesopotamians invented, the teacher accepted children’s answer that ‘cities’ were invented by Mesopotamians.

A teacher explained that irrigation was ‘digging channels’. This is not correct as irrigation simply means bringing water artificially to the land by means of ditches or channels. Another teacher explained ditches as ‘lines’ in the ground. Children were confused as to how these lines could carry water.

It has to be mentioned that all the teachers observed were qualified trained teachers, and in every school we visited we also observed the literacy leads who were experienced teachers.

### Departure from crucial aspects of the protocol

In a few lessons observed key aspects of the protocol were not adhered to. For example, instead of reading the text aloud to the class one teacher appointed pupils to read the text. Another teacher thought the keyword exercises, which were crucial in building vocabulary, were optional and so the children missed most of these exercises. In another school, one teacher did not mark and correct mistakes in pupils' exercise books.

One important aspect of the curriculum was immediate feedback. Children's exercises were supposed to be marked in the lesson and mistakes pointed out to them and corrected, as soon as possible. In a few sample exercise books we looked at, a number of grammatical and spelling mistakes in pupils' workbooks were not picked up. Pupils were also supposed to use the keywords in different contexts. However, in most cases pupils simply used the keywords but not in a way that demonstrates that they have understood the word. It was quite common to see pupils simply repeating the word but in a question form or rephrasing the statement. Using the word 'democracy' as an illustration, sentences children made could be:

"What is democracy?" or

"My teacher asked us to make a sentence using the word democracy."

What was surprising was that none of the teachers insisted that pupils used the keywords to demonstrate understanding. Although there was a lot of repetition within a lesson, there was no attempt to experiment with the keywords or vocabulary taught, such as using the keywords in different contexts.

## **Implications and conclusions**

### *Substantive*

The impact evaluation shows no discernible effect of the WWR programme on children's reading. This suggests that simply asking teachers to teach factual knowledge does not work in practice in raising children's literacy. However, the process evaluation suggests that the programme has benefited children in some ways. For example, children are answering in complete sentences, using sentence starters and also thinking more about the keywords. Therefore, this approach could be used if the purpose was to teach more about history and geography with no discernible damage to literacy. The lack of measurable impact on literacy itself could be due to a number of factors.

### Facts or thinking skills?

First, there is no evidence, as far as we know, that the Core Knowledge Sequence as originally conceived by E.D. Hirsch has been effective in improving the reading skills of children. The evidence so far has been patchy, inconsistent and mixed. And even if we assume that it does, the Word and World Reading programme being evaluated in the pilot trial differs from the Core Knowledge Sequence in important ways.

The original idea of the Core Knowledge Sequence was to teach facts to pupils to give them the foundation to think critically and opportunities to apply the knowledge and

to question the facts. The programme as used in some US schools encouraged pupils to think with rich questioning and discussion topics. Hirsch recommended that 50% of the curriculum be devoted to the teaching of content and the other 50% for wider discussions of the topics covered. The WWR programme, on the other hand, was very prescriptive and the lessons were scripted. There was little interaction and opportunities for in-depth discussions as teachers were given set questions and set answers. Pupils were not encouraged to question or explore the content, or to think critically.

However, the point of education is not simply the acquisition of knowledge, but also skills to synthesise and comprehend the information confronted. This is especially so in the new global knowledge economy where children are constantly bombarded with information. The profusion of information- generating devices like the mobile phones, iPhones, iPads, Blackberry, and modern means of sharing and accessing information like google, twitter, blogs and instant messaging make it easy for children to access information. Therefore, what children need is not simply more information, but the ability to sieve through the information, to be able to judge what is believable and what is not, to evaluate the evidence, to interpret the data received and reported and to critically appraise the quality of such information. It is a useful skill for young people to be able to evaluate the integrity and validity of information they are confronted with, weigh the evidence presented to them, and make judgements about them. It is through this process that knowledge can be developed and verified. Such skills are therefore even more necessary in the 21st century (but none of this was included by TCC in their intervention). This trial suggests the need for both core knowledge and thinking and processing skills.

#### Duration

The WWR programme also differs from the original Core Knowledge Sequence in that it was funded for and lasted only one year. Almost all the evaluations of the Core Knowledge Sequence were for three years. This was to allow for children to build and develop key concepts and vocabulary. The theory is that children's learning of factual knowledge needs to be built up over time. The one-year pilot, as devised by the developers, is perhaps therefore not long enough for the effects to be apparent. As the theory suggests, it is possible that the concepts and knowledge learned could remain latent, so the benefits of the programme may not be realized in the short term. There may be long-term gains as children's latent knowledge accumulates and feeds on to learning in other areas. This benefit has not been tested. Perhaps future research could look into the longitudinal effects of the Core Knowledge curriculum, comparing children exposed to the curriculum with those who were not.

It may also be beneficial if children start early for longer exposure to the programme. The strongest evaluations of the Core Knowledge curriculum have shown that positive effects were found among younger pre-school children. Results for older children (grade one to five) were generally less convincing.

#### Poor implementation of the intervention

Formative feedback from the process evaluation suggests that there are aspects of the programme that could be improved.

Choice of vocabulary, concepts and activities should be carefully considered for their suitability for children of different ages and ability.

Since the programme was specifically designed for primary school aged children, the textbooks could be presented more attractively. For example, the front cover page could be made more attractive to encourage interest in the subjects. Difficult concepts could be presented in pictorial forms or even in cartoons. Pupils also suggested fun activities like puzzles, crosswords and word search.

The programme offered little opportunity for teachers to be creative. Only a handful of teachers attempted to be creative by supplementing the lessons with their own video clips and slides. The programme could include a list of suggested activities for teachers to use with the pupils. For example, teachers could use Google Maps and satellite images to teach directions, continents and oceans. Using something that children are familiar with can make the lessons more interesting and relevant to them.

Due to the prescriptive nature of the programme, there was little scope for discussion. One recommendation, therefore, is to scale down the number of topics to allow teachers more time to cover each one in greater depth. These topics could be developed as the child moves up each level, progressing from simple concepts to more complex ideas. The children interviewed commented that they did not have the chance to discuss, and suggested that they would rather have fewer topics to allow them “to explore the subjects in-depth”. In a lesson on ancient civilisations the children asked how we knew what Mesopotamia was like. So an introductory lesson on sources used in Geography or History may be beneficial.

The one-year intervention was perhaps too short for children to build and develop their vocabulary, and could be extended to three years. In line with the Core Knowledge theory of learning, the vocabulary should be built up over time. The short duration of the pilot trial did not allow this to happen.

There were some excellent lessons, although these were a minority. More opportunities could be created for teachers to observe each other's lessons, share experiences, resources and tips about how to make the lessons interesting. Schools could organise staff workshops for CPD (continuing professional development) training where they share and produce teaching materials. Teachers expressed that they would relish the idea of sharing ideas and resources.

Teaching content knowledge requires different skills to teaching language. It requires some basic background knowledge of the subjects too. Therefore to use English language teachers who are trained to teach English to teach content subjects would require proper prior training. The training that teachers received in this pilot focused on the theory of learning and the use of the Core Knowledge Sequence. Intensive professional training of teachers is therefore necessary. Such training should cover both the contents of History and Geography used in the curriculum and also relevant pedagogical skills required for teaching geographical and historical concepts.

Perhaps teachers as professionals do not respond well to such a prescriptive protocol. It would be interesting to find out if it would be more effective if teachers were given

the same texts and content, but allowed to use their own resources and skills to teach the topics.

### Summary

In summary, the findings of the study suggest that simply having a prescriptive curriculum to teach facts or knowledge alone is not enough. It cannot overcome teachers' lack of content knowledge.

There are other complexities involved that need to be considered, such as the pedagogical skills of teachers, availability of appropriate resources and the fact that teachers as professionals may not respond well to being dictated on how to teach and what to teach. However, for those who support the notion that increasing the world knowledge of young children in History and Geography is a good thing in itself, and ought not necessarily lead to immediate improvements in literacy, the findings of this small trial might be reassuring. At least, teaching core knowledge does not appear to harm literacy.

### **Acknowledgements**

This project was funded by the Education Endowment Foundation. We are grateful to the staff and students of the schools agreeing to take part.

### **References**

- Berk, R. and Freedman, D. (2001) Statistical assumptions as empirical commitments, <http://www.stat.berkeley.edu/~census/berk2.pdf>, accessed 03/07/14
- Baer, J. (2011) The impact of the Core Knowledge Curriculum on creativity, *Creativity Research Journal*, 15, 2/3, 297-300.
- Booth, W. C. (1988) Cultural literacy and liberal learning: An open letter to E. D. Hirsch Jr., *Change* 20, 4, 11-21.
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J. and Balain, S. (2007) A conceptual framework for implementation fidelity, *Implementation Science*, 2, 1, 40.
- Carver, R. (1978) The case against statistical significance testing, *Harvard Educational Review*, 48, 378-399
- Christodoulou, D. (2014) *Seven myths about education*, London: Routledge
- Core Knowledge Foundation (2000) *In Oklahoma City, a rigorous scientific study shows the positive equity effects of Core Knowledge*, [http://coreknowledge.org/CK/about/research/oklahoma\\_study\\_2000.htm](http://coreknowledge.org/CK/about/research/oklahoma_study_2000.htm), accessed April 24 2009
- Core Knowledge (2014) *E.D. Hirsch and Cultural Literacy*, <http://www.coreknowledge.org.uk/culturalliteracyck.php>
- Coughlan, S. (2013) Gove sets out 'core knowledge' curriculum plans. *BBC News*, 6 February 2013. Accessed 19 November 2015, <http://www.bbc.co.uk/news/education-21346812>
- Dartnow, A., Borman, G. and Springfield, S. (2000) School reform through a highly specified curriculum: implementation and effects of the Core Knowledge Sequence, *The Elementary School Journal*, 101, 2, 167-192

- Falk, R. and Greenbaum, C. (1995) Significance tests die hard: the amazing persistence of a probabilistic misconception, *Theory and Psychology*, 5, 75-98
- Gibb, N. (2014) The importance of teaching. Speech given at ResearchEd conference in London. <https://www.gov.uk/government/speeches/the-importance-of-teaching>, accessed February 2015.
- Gorard, S., See, B.H. and Morris, B. (2016) *Identifying the most effective teaching pedagogies in primary schools*. London: DfE.
- Gorard, S. (2016) Damaging real lives through obstinacy: re-emphasising why significance testing is wrong, *Sociological Research On-line*, (forthcoming)
- Gorard, S. and Gorard, J. (2016) What to do instead of significance testing? Calculating the 'number of counterfactual cases needed to disturb a finding', *International Journal of Social Research Methodology*, <http://tandfonline.com/doi/full/10.1080/13645579.2015.1091235>
- Gorard, S., Siddiqui, N. and See, B.H. (2017) Can 'Philosophy for Children' improve primary school attainment, *Journal of Philosophy of Education*, (forthcoming)
- Hirsch, E. (1987) *Cultural literacy: what every American needs to know*, New York: Vintage
- Iver, M., Stringfield, S. and McHugh, B. (2000) *Core Knowledge Curriculum: Five-year analysis of implementation and effects in five Maryland Schools*. Report No. 50. Baltimore: John Hopkins University
- Lipsey, M., Puzio, K., Yun, C., Hebert, M., Steinka-Fry, K., Cole, M., Roberts, M., Anthony, K. and Busick, M. (2012) Translating the statistical representation of the effects of education interventions into more readily interpretable forms, Washington DC: Institute of Education Sciences
- Mitra, S. (2016) *Schools should scrap the 3Rs*. TES, 20<sup>th</sup> January 2016. <https://www.tes.com/news/school-news/breaking-news/sugata-mitra-schools-should-scrap-3rs>. Accessed on 23 January 2016.
- Montgomery, P., Underhill, K., Gardner, F., Operario, D. and Mayo-Wilson, E. (2013) The Oxford Implementation Index: a new tool for incorporating implementation data into systematic reviews and meta-analyses, *Journal of Clinical Epidemiology*, 66, 8, 874- 882
- Schwarz, G. (1988) A good idea gone wrong, *Educational Leadership*, 45, 4, 77.
- Squire, J. (1988) Basic skills are not enough, *Educational Leadership*, 45, 4, 76-77.
- Slavin, B. (2016) Farewell to the walking encyclopaedia. *Huffpost Education*. 28 January 2016. [http://www.huffingtonpost.com/robert-e-slavin/farewell-to-the-walking-e\\_b\\_9100144.html](http://www.huffingtonpost.com/robert-e-slavin/farewell-to-the-walking-e_b_9100144.html). Accessed on 29 January 2016.
- Steckler, A. and Linnan, L. (2002) *Process evaluation for public health interventions and research*, San Francisco: Jossey-Bass.
- Stringfield, S., Dartnow, A., Borman, G. and Rachuba, L. (2000) *National evaluation of Core Knowledge Sequence implementation Final Report*. Report No. 49. Baltimore: CRESPE, John Hopkins University
- Tchudi, S. (1988) Slogans indeed: A reply to Hirsch, *Educational Leadership*, 45, 94, 72-74.
- The Guardian (2012) US idea of 'cultural literacy' and key facts child should know arrives in UK. <http://www.theguardian.com/education/2012/oct/15/hirsch-core-knowledge-curriculum-review>. Accessed 28 October 2015
- The NYC Core Knowledge Early Literacy Pilot (2012) [http://www.coreknowledge.org/mimik/mimik\\_uploads/documents/712/CK%20Early%20Literacy%20Pilot%203%2012%2012.pdf](http://www.coreknowledge.org/mimik/mimik_uploads/documents/712/CK%20Early%20Literacy%20Pilot%203%2012%2012.pdf), accessed July 2014

- Watts, D. (1991) Why is introductory statistics difficult to learn?, *The American Statistician*, 45, 4, 290-291
- Whitehurst-Hall, J. (1999) *The impact of the Core Knowledge Curriculum on the achievement of seventh and eighth grade students*, PhD dissertation, Faculty of Education, University of Georgia, Athens
- Young, T. (2012) Free schools: The research lab of state education. *The Guardian*, 3 October 2012. <http://www.theguardian.com/teacher-network/2012/oct/03/free-schools-research-development-education>. Accessed 28 October 2015